

Códigos consonánticos en 94 lenguas austroasiáticas

Doce ramas repartidas: la familia que separa el largo de la palabra del tamaño de la muestra

Alejandro Toledo Martínez · ORCID 0009-0000-1277-9697 · Corpus Integrativo · 2026

Resumen

Presentamos un catálogo de 2.135 códigos consonánticos que agrupan 14.851 formas léxicas de 94 lenguas austroasiáticas, construido sin etimología, protoformas ni juicios de cognación.

Es la primera familia de la serie donde los códigos de mayor alcance no son de una sola consonante. $m \cdot t$ para *ojo* aparece en 64 lenguas y once de las doce ramas; $m \cdot h$ para *nariz*, en 58 y diez. Son el *mat* y el *mu* austroasiáticos, los dos comparanda que la comparación de esta familia lleva un siglo citando, y el procedimiento no los buscó: salieron de exigir que un esqueleto se repita para el mismo concepto en tres lenguas.

Y esta familia falsa un hallazgo de la serie. El paper sino-tibetano registró y confirmó fuera de muestra que el nulo del agrupamiento se dispara por debajo de 2,8 consonantes por palabra. El austroasiático tiene 2,75 consonantes y un nulo del 14,6 %; el austronesio tiene 2,78 y un nulo del 39,0 %. Misma longitud de palabra, veinticuatro puntos de diferencia. Lo que los separa son las lenguas: 94 frente a 582.

Sobre las nueve familias medidas, $r(\text{aridad, nulo}) = -0,84$ y $r(\log \text{ lenguas, nulo}) = +0,70$, y las dos variables son en buena parte independientes entre sí. El nulo es función del largo de la palabra Y del tamaño de la muestra, y las dos estaban confundidas porque hasta ahora toda familia de palabra corta era además grande.

También se documenta un bug encontrado antes de construir el catálogo —la cuarta vez que un no-consonante entra en el esqueleto por una convención de la fuente, y la primera que se caza a tiempo.

1. Qué se busca, y por qué esta familia

El procedimiento está descrito una sola vez para toda la serie en «El catálogo de códigos consonánticos» y no se repite. Lo imprescindible: un código es un esqueleto consonántico que aparece para el mismo concepto en tres lenguas distintas o más, y no se afirma que sus miembros sean cognados, ni que descendan de una protoforma, ni que el código signifique nada.

Esta familia estaba en el plan desde el principio y se saltó dos veces. Iba tercera —«doce ramas equilibradas y tono en el 30 % de las formas: es el punto medio exacto del eje»— y era donde iba a decidirse qué hacer con el tono «en una familia donde el resto de la estructura es favorable». El tono acabó decidiéndose sobre la marcha en sino-tibetano, que es justo lo que el plan quería evitar.

Llega en noveno lugar por otra razón, y es mejor. Con ocho familias hechas, el criterio para elegir la siguiente dejó de ser «ramas equilibradas» y pasó a ser ramas repartidas $\times$ palabra larga: en familias de palabra corta el cruce de ramas se va a los códigos unarios, donde el catálogo no puede distinguir un etimón monoconsonántico verdadero de una coincidencia por disponibilidad. El austroasiático es la única pendiente que tenía las dos cosas.

2. La población

94 lenguas · 49.783 formas · 1.570 conceptos · 12 ramas.

Rama	Lenguas	Formas
Bahnárico	23	6.291
Aslian	15	5.694
Viético	13	7.221
Khasi-Palaung	13	13.469
Mundaico	8	753
Khmuico	6	5.424
Katuico · Peárico	4 + 4	5.564
Mángico	3	3.454
Nicobarés · Khmérico · Mónico	2+2+2	1.913

100 % de las formas obtienen esqueleto, y ninguna lengua se excluye. La rama mayor es el 24 % —el mejor reparto de la serie después del pama-ñungano (18 %)—, y nueve de las doce ramas tienen tres lenguas o más.

Una sola macroárea: la lista de anomalías del §6.1-bis no existe aquí, y se dice antes de los resultados.

2.1 Ninguna lengua se excluye, y las tres propuestas se rechazan

Dos pares de duplicados por contenido, los dos por debajo de la regla. **KNB** (a Pearic variety) y **Surin Khmer** coinciden en el 65 % de 936 conceptos; **Kensiw Kedah** y **Kintaq**, en el 61 % de 145. La regla publicada exige el 80 % y sus glotocódigos difieren. Son vecinas de la misma rama, que es lo que se espera que pase.

Korku está a 2.845 km del centro, 4,9 veces la mediana. Es mundaica, hablada en Madhya Pradesh, y la familia se extiende de la India a Vietnam. Es la quinta vez que este chequeo pide de más, y la regla ya corregida se aplica: la geografía sola no propone, acompaña.

Quince lenguas marcan tono en dígitos donde el resto no lo hace, casi todas viéticas. Es una propiedad de la fuente y no de la lengua, y el esqueleto ignora el tono en todo caso.

2.2 Un bug encontrado ANTES de construir el catálogo

Cuatro lenguas peáricas —chong, samre, kasong, **KNB**—, el **wa** y siete más escriben el tono de registro como una letra tras punto: **pi?**.A, **klaŋ**.B, **krih**.C. Es la convención de Diffloth y Ferlus, y es tono: no es un segmento.

La **A** es vocal y se caía sola. La **B**, la **C** y la **D** son consonantes y entraban en el esqueleto:

Wa **klaŋ**.A_ćε.B → k·l·n·s·b ← la «b» final era el tono B
 Wa **pau?**.B_?aik.A_tañ.A → p·h·b·h·k·t·n ← la «b» de la tercera posición era el tono B

12.974 formas en 12 lenguas llevan la marca; 5.866 la tenían como segmento —en las otras la fuente ya la había convertido a superíndice— y 3.235 acababan con una consonante inventada dentro del esqueleto.

Es la cuarta vez que un no-consonante entra en el esqueleto por una convención de la fuente, después del *stød* danés, la *ø* sueca leída como *θ* y los dígrafos prenasalizados. Las tres anteriores se encontraron *después* de publicar la familia afectada. Ésta la encontró la fase 0 antes de construir el catálogo, que es exactamente para lo que existe esa fase, y es la primera vez que el orden sale bien.

El arreglo es conservador, como los otros tres: alinea las palabras de la grafía con los grupos de segmentos y solo quita el último segmento de un grupo cuando su palabra termina en .X y ese segmento es exactamente esa letra en minúscula. Una b final de verdad no se pierde. Comprobado sobre las 5.866 formas: cero discrepancias entre el esqueleto guardado y el recalculado.

3. Qué cuesta el instrumento aquí

Aridad media: 2,75 consonantes por forma. Colisión: 56,2 %. Más de la mitad de las formas comparten esqueleto con otro concepto dentro de su propia lengua.

Las dos cifras la sitúan en el lado caro del eje, entre el austronesio (2,78) y el trans-neoguineano (2,41). Lo que §5 mide es que ahí no acaba la historia.

El tono no entra en el esqueleto, y treinta lenguas lo marcan de una u otra forma. Se declara como dimensión que el esqueleto no captura, igual que las vocales.

Cinco conceptos están duplicados por caja en la lista de origen —HAIR/hair, KNOW/know, SMELL/smell, THINK/think, HIDE/hide—, así que el mismo esqueleto puede contar dos veces. Es la colisión de la lista que el paper trans-neoguineano mide sobre las formas, aquí en su forma más cruda: la misma glosa escrita de dos maneras.

4. El catálogo

2.135 códigos sobre 1.570 conceptos, que agrupan 14.851 formas de 49.783 (29,8 %). De ellos, 1.695 son binarios o ternarios y 309 son unarios: el 14,5 %.

Aridad	Códigos	%
1	309	14,5 %
2	1.243	58,2 %
3	452	21,2 %
4	114	5,3 %
5+	17	0,8 %

El 14,5 % de unarios es el segundo más bajo de la serie, tras el 3,9 % del pama-ñungano y muy lejos del 35,9 % sino-tibetano. Y el 58,2 % de binarios es el más alto: esta familia concentra su catálogo justo en la aridad que el instrumento maneja mejor.

Precisión: 6,63 % frente a 0,144 % del azar = $46 \times [45 - 46 \times \text{según semilla}]$.

Y sobre el agrupamiento sin umbral de tres lenguas: 41,7 % real frente a 14,6 % del nulo — ventaja de 27,1 puntos, $2,9 \times$ el azar, $\kappa = 0,317$.

4.1 Los códigos de mayor alcance son binarios, y cruzan once ramas

Los quince de más alcance, sin filtrar:

Código	Concepto	Lenguas	Aridad	Ramas	Enriq.
m·t	<i>ojo</i>	64	2	11	147×
m·h	<i>nariz</i>	58	2	10	100×
k	<i>pez</i>	44	1	8	42×
t	<i>mano</i>	44	1	11	49×
s·m	<i>pájaro</i>	42	2	8	93×
s·k	<i>pelo</i> (HAIR)	40	2	8	82×
c·n	<i>pie</i>	37	2	8	61×
s·n	<i>pie</i>	37	2	9	33×
k·p	<i>morder</i>	36	2	5	83×
s	<i>perro</i>	36	1	8	32×
p·r	<i>volar</i>	35	2	9	143×
m	<i>tú</i>	32	1	8	36×
t·n	<i>tejer</i>	32	2	8	78×
m·j	<i>uno</i>	31	2	7	153×
t·h	<i>pecho</i>	30	2	7	45×

Once de los quince son binarios, y los dos primeros lo son. Es la primera vez en la serie: en sino-tibetano los doce primeros eran unarios, en trans-neoguineano catorce de quince, en austronesio y pama-ñungano el techo también estaba tomado por códigos de una sola consonante.

m·t|ojo es el *mat* austroasiático y **m·h|nariz** el *muh*, los dos comparanda de manual. Sus repartos:

Código	Concepto	Lenguas	Ramas	Las cinco mayores
m·t	ojo	64	11	Bahnárico 21 · Aslian 15 · Viético 13 · Khmuico 4 · Peárico 3
m·h	nariz	58	10	Bahnárico 23 · Aslian 14 · Khmuico 5 · Khasi-Palaung 4 · Katuico 4

No están concentrados en una rama con cola: cubren la familia. El procedimiento no consultó ninguna reconstrucción.

5. El hallazgo: el nulo no es función de la aridad sola

El paper sino-tibetano publicó, sobre siete familias y confirmado fuera de muestra en la octava, que el nulo del agrupamiento se dispara por debajo de tres consonantes por palabra. La predicción se registró antes de calcular y se cumplió: aridad 2,22, nulo 42,0 %.

El austroasiático la rompe.

Familia	Aridad	Lenguas	Nulo	κ
Sino-tibetano	2,22	267	42,0 %	0,340
Trans-neoguineano	2,41	196	37,4 %	0,232
Austroasiático	2,75	94	14,6 %	0,317
Austronesio	2,78	582	39,0 %	0,380
Turco	2,85	32	13,7 %	0,498
Afroasiático	3,16	111	4,7 %	0,262
Urálico	3,27	27	5,0 %	0,301
Indoeuropeo	3,35	207	11,8 %	0,364
Pama-ñungano	3,35	154	11,3 %	0,310

Léase la tercera y la cuarta fila juntas. Austroasiático y austronesio tienen prácticamente la misma longitud de palabra —2,75 y 2,78— y sus nulos difieren en veinticuatro puntos. Lo que los separa es el número de lenguas: 94 contra 582.

Y no es un caso aislado. Sobre las nueve:

	r
aridad $\leftrightarrow$ nulo	−0,84
log lenguas $\leftrightarrow$ nulo	+0,70
aridad $\leftrightarrow$ log lenguas	−0,34

Las dos variables predicen el nulo y son en buena parte independientes entre sí. La correlación entre ellas es débil, así que no se están midiendo dos veces lo mismo.

5.1 Por qué, y por qué estaba escondido

El mecanismo no tiene misterio. El nulo pregunta con qué frecuencia dos lenguas cualesquiera comparten esqueleto para el mismo concepto al barajar. Con 582 lenguas hay 169.071 pares donde eso puede pasar; con 94, 4.371. Más lenguas es más oportunidades de coincidir por azar, exactamente igual que más consonantes es menos.

Lo que lo mantenía escondido es un accidente del orden de trabajo: las tres familias de palabra corta que la serie había hecho eran también las grandes —sino-tibetano 267, trans-neoguineano 196, austronesio 582—, y las de palabra larga incluían las dos más pequeñas —turco 32, urálico 27—. Aridad y tamaño venían correlacionados en la muestra sin estarlo en el mundo.

El austroasiático es la primera familia de palabra corta y muestra pequeña, y por eso separa las dos.

Y no es el número de conceptos. Es la objeción obvia —un corpus con más conceptos reparte las formas en más cubos y baja la coincidencia al barajar— y no se sostiene: $r(\log \text{ conceptos}, \text{ nulo}) = -0,09$ sobre las nueve. El caso que decide es el mismo: austroasiático y austronesio tienen 1.570 y 1.823 conceptos, un 16 % de diferencia, y nulos de 14,6 % y 39,0 %.

Merece anotarse porque no era predecible: el azar de la *precisión* sí depende fuertemente del número de conceptos —el paper indoeuropeo lo publica— y cabría esperar que el nulo del agrupamiento se comportara igual. Son dos nulos de dos preguntas distintas, y responden a variables distintas.

5.2 Qué queda en pie del hallazgo anterior

Casi todo, con una condición añadida. La aridad sigue siendo el predictor más fuerte (−0,84 contra +0,70), y el escalón por debajo de 2,8 sigue siendo real *a igualdad de tamaño*. Lo que no se sostiene es leerlo como una función de una sola variable, y en particular:

Una cifra bruta de agrupamiento no se puede comparar entre familias sin su nulo al lado, y el nulo no se puede estimar solo desde la aridad. El paper trans-neoguineano formuló la primera mitad de esa regla; ésta es la segunda.

Y κ no sigue a ninguna de las dos. El austroasiático da 0,317 con nulo bajo; el sino-tibetano, 0,340 con nulo altísimo. Coste y rendimiento son cosas distintas, que es lo que §5 del paper sino-tibetano ya había separado.

6. Lo que la familia da

6.1 El cruce de ramas, y por primera vez sin que se lo lleven los unarios

641 códigos cruzan tres ramas o más, y 233 cruzan cuatro. Descompuesto por aridad, que es la lectura que el paper sino-tibetano obligó a publicar:

Aridad	Códigos	Cruzan ≥ 3 ramas	Tasa
1 (unario)	309	143	46,3 %
2	1.243	413	33,2 %
3	452	65	14,4 %
4+	131	20	15,3 %

Solo el 22 % de los que cruzan son unarios, contra el 58 % del sino-tibetano y el 64 % del trans-neoguineano. La cifra defendible es 478 —los binarios y ternarios—, y es la segunda mejor de la serie tras los 1.237 del indoeuropeo, con menos de la mitad de lenguas que él.

La tasa sigue cayendo con la aridad, como en todas las familias: un código corto cruza más fácil. Pero aquí la caída es de 46 % a 33 %, no de 28 % a 3 %, porque hay ramas de verdad que cruzar.

6.2 Pares de sustitución: 631 candidatos disjuntos

De 754 pares de códigos del mismo concepto que difieren en una sola posición, 631 son disjuntos en lenguas y solo esos pueden proponerse como isoglosa. El 16 % que solapa es la segunda proporción más limpia de la serie, tras el 5 % del pama-ñungano.

Los dos códigos para «pie» — $c \cdot n$ en 37 lenguas y ocho ramas, $s \cdot n$ en 37 y nueve— son el mejor ejemplo de lo que esta sección produce: mismo concepto, misma segunda consonante, y la primera reparte la familia.

6.3 Cuatro clases emergentes, y es la tercera familia de nueve que produce alguna

Par	Veces (descubre)	Razón	Veces (comprueba)	Razón
$b \leftrightarrow \mathfrak{b}$	183	45,7×	91	21,1×
$c \leftrightarrow s$	3.345	19,5×	3.115	16,8×
$p \leftrightarrow \mathfrak{p}$	106	9,1×	172	13,8×
$s \leftrightarrow z$	42	4,8×	87	9,2×
$d \leftrightarrow t$	931	5,3×	1.549	8,1×
$b \leftrightarrow p$	657	5,4×	828	6,3×
$c \leftrightarrow j$	299	2,8×	647	5,7×

El cierre a $\geq 2\times$ da cuatro clases limpias:

CE1 { c d h j s t z } CE2 { b p $\mathfrak{b}$ }
CE3 { l r x } CE4 { g k }

Quince segmentos en cuatro clases, y las cuatro son reconocibles: coranales, labiales, líquidas y velares. Solo el austronesio y el pama-ñungano habían producido clases limpias; las otras seis familias de la serie se publican sin ninguna porque el cierre les mete casi todo el inventario en una sola.

$b \leftrightarrow \mathfrak{b}$ a 21× y $p \leftrightarrow \mathfrak{p}$ a 13,8× son la implosiva bilabial frente a sus dos correspondientes pulmonares — una alternancia real de esta zona lingüística, que el procedimiento encuentra sin saber qué es una

implosiva. No son correspondencias regulares: falta entorno y falta dirección. Son candidatas, con el sitio donde buscarlas.

6.4 La geografía, con una inversión que hay que declarar

Códigos compartidos	Pares	Distancia media	vs azar
1	1.485	1.243 km	1,07
2	486	1.475 km	1,26
3–5	1.135	1.135 km	0,97
6–10	790	887 km	0,76
11–20	317	581 km	0,50
21+	416	288 km	0,25

Baja a 0,25 del azar, la segunda mejor de la serie tras el 0,09 del trans-neoguineano. Pero el segundo cubo va al revés —1,26, más lejos que el azar— y eso se publica en vez de suavizarse: con solo 486 pares es el cubo más pequeño de la tabla, y compartir exactamente dos códigos es casi indistinguible de compartir uno. La monotonía se recupera desde el tercero y es limpia a partir de ahí.

7. Límites

Una sola macroárea: el §6.1-bis no existe.

El tono no entra en el esqueleto, y treinta lenguas lo marcan. En las que lo tienen como contraste, el esqueleto ve menos de lo que la lengua distingue.

El 14,5 % del catálogo es de una sola consonante, y el catálogo no puede distinguir el etimón monoconsonántico verdadero de la coincidencia. Es la segunda proporción más baja de la serie y sigue siendo una de cada siete: los unarios se cuentan y no se leen.

El bahnárico es el 24 % de las lenguas. Es el segundo mejor reparto de la serie y sigue siendo una rama que pesa más que las demás — y pesa especialmente en el techo de §4.1, donde encabeza los siete primeros códigos.

Cinco conceptos están duplicados por caja en la lista de origen, así que cinco esqueletos pueden contar dos veces.

El despiece automático no se usa, como en las ocho anteriores. Aquí cuesta $\Delta\kappa = -0,005$, la segunda caída más pequeña de la serie — y merece decirse que ocurre en una familia célebre por su infijación, donde cabría esperar que un descubridor de afijos tuviera más que ganar. No lo gana.

8. Conclusión

El plan de la serie ponía a esta familia en tercer lugar y se saltó dos veces. Llega la novena, y llega mejor de lo que se esperaba por una razón que el propio plan tuvo que corregir antes: lo que hace productiva a una familia no es que tenga ramas repartidas, sino que las tenga Y tenga palabras suficientemente largas. El austroasiático es la única pendiente que tenía las dos, y es la primera de la serie cuyo techo por alcance no lo toman los códigos unarios: $m \cdot t$ para *ojo* en once ramas de doce, $m \cdot h$ para *nariz* en diez, sin haber consultado ninguna reconstrucción.

Y el resultado de método viene de romper algo. El nulo del agrupamiento se venía leyendo como una propiedad del largo de la palabra, con una predicción registrada y confirmada que lo sostenía. Esta

familia lo parte en dos: con la misma aridad que el austronesio y una sexta parte de sus lenguas, su nulo es veinticuatro puntos menor. El largo de la palabra y el tamaño de la muestra son dos ejes, no uno, y venían confundidos porque en las ocho familias anteriores iban juntos por accidente del orden de trabajo.

Ninguna de las dos cosas se habría visto sin hacer la novena. Y la que más cuesta admitir es la segunda: el hallazgo que esta serie registró de antemano, confirmó fuera de muestra y publicó como su regularidad más sólida, era la mitad de una regularidad.

Referencias

- Diffloth, G. (2005). The contribution of linguistic palaeontology to the homeland of Austroasiatic. En *The Peopling of East Asia*. RoutledgeCurzon. — [el tono de registro escrito como letra, §2.2]
- Ferlus, M. (1979). Formation des registres et mutations consonantiques dans les langues môn-khmer. *Mon-Khmer Studies* 8.
- Hammarström, H., Forkel, R., Haspelmath, M. y Bank, S. *Glottolog*. Max Planck Institute for Evolutionary Anthropology.
- Hubert, L. y Arabie, P. (1985). Comparing partitions. *Journal of Classification* 2(1). — [la corrección por azar que este paper usa como κ]
- Jäger, G. (2018). Global-scale phylogenetic linguistic inference from lexical resources. *Scientific Data* 5.
- List, J.-M. (2014). *Sequence Comparison in Historical Linguistics*. Düsseldorf University Press.
- List, J.-M., Forkel, R., Greenhill, S. J. et al. (2022). Lexibank, a public repository of standardized wordlists. *Scientific Data* 9.
- List, J.-M., Cysouw, M. y Forkel, R. (2016). Concepticon: a resource for the linking of concept lists. *LREC*.
- Shorto, H. (2006). *A Mon-Khmer Comparative Dictionary*. Pacific Linguistics. — [los comparanda de §4.1, que este trabajo no consultó]
- Sidwell, P. y Blench, R. (2011). The Austroasiatic Urheimat: the Southeastern Riverine Hypothesis. En *Dynamics of Human Diversity*. Pacific Linguistics.
- Toledo Martínez, A. (2026). *El catálogo de códigos consonánticos: un método comparativo sin genealogía, cognación ni reconstrucción*. — [el método de la serie, descrito una sola vez]
- Toledo Martínez, A. (2026). *Códigos consonánticos en 267 lenguas sino-tibetanas*. — [el hallazgo del escalón de aridad que este paper parte en dos]
- Toledo Martínez, A. (2026). *Corpus Integrativo: una base léxica multicapa de 225 macrosistemas*. — [los datos] · <https://doi.org/10.13140/RG.2.2.31244.68486>

Datos y reproducibilidad

Catálogo	familias/austrasiatic/output/catalogo-austrasiatic.tsv — 2.135 filas
Detalle por forma	...-lenguas.tsv — 13.092 filas

Higiene y decisiones (§2.1)	pipeline/fases/fase0_higiene.py → runs/ fase0-decisiones.md
Bug del tono en letra (§2.2)	corpus_integrativo/ingest/recompute_tono_letra.py · respaldo skeleton_bak_tono letra_20260909
Tabla de normalización (§3)	pipeline/fases/tabla_normalizacion.py → runs/ normalizacion.md
Catálogo (§4)	pipeline/fases/catalogo_familia.py
Agrupamiento y nulos (§4, §5)	pipeline/fases/agrupamiento_familia.py → runs/ agrupamiento.md
La tabla de las nueve (§5)	pipeline/tools/tabla_serie.py — derivada, no escrita a mano
Despiece automático (§7)	pipeline/fases/fase2_despiece.py → runs/fase2-despiece. md
Cruce de ramas (§6.1)	pipeline/fases/fase6_por_rama.py → runs/fase6-por-rama. md
Códigos hermanos (§6.2)	pipeline/fases/codigos_hermanos.py → runs/ codigos-hermanos.md
Clases emergentes (§6.3)	pipeline/fases/fase6_clases_emergentes.py → runs/ fase6-clases.md
Geografía (§6.4)	pipeline/fases/codigos_vs_parentesco.py → runs/ codigos-parentesco.md

Toda cifra sale de esos scripts sobre la misma población declarada en §2, con las mismas exclusiones —que aquí son ninguna—. Ninguna se escribió a mano.

Anexo A · Cien códigos comentados

El catálogo tiene 2.135 códigos y los datos completos están en el directorio output/ de la familia. Lo que sigue es una muestra legible: los cien códigos binarios y ternarios de mayor alcance, cada uno con hasta dieciséis lenguas y sus palabras.

Tres cosas sobre la presentación. Las lenguas están repartidas a lo largo de la lista alfabética, no tomadas de su cabeza. Los nombres largos se recortan; las formas no se alteran. Y la agrupación temática es un orden de lectura declarado a mano, no una clasificación semántica.

Este anexo sí lleva columna de código emergente, y es la primera vez desde el pama-ñungano. El cierre de la fase 6 da aquí cuatro clases limpias —{c d h j s t z}, {b p ɓ}, {l r x}, {g k}— y esta familia es la tercera de nueve que produce alguna. En las seis restantes el cierre mete casi todo el inventario en una sola clase y el anexo se publica sin columna.

La columna dice lo que dice y nada más: dos códigos con el mismo dígito difieren en segmentos que la fase 6 encontró intercambiables *en esta familia*. No son correspondencias regulares —falta entorno y falta dirección— y el valor de una clase es relativo a la familia que la produjo.

Qué selecciona este anexo, medido

Ordenar por alcance no es neutral. Medido con `pipeline/tools/audita_anexo.py`:

	catálogo	este anexo
códigos binarios	58,2 %	84 %
códigos ternarios	21,2 %	16 %
de rama dominante bahnárica	21,3 %	58 %
de rama dominante khasi-palaung	24,4 %	11 %
de rama dominante peárica	16,6 %	0 %
que cruzan cuatro ramas o más	10,9 %	60 %
enriquecimiento medio	200×	131×

Es el anexo menos deformado de la serie por el lado de la aridez: 84/16 contra el 58/21 del catálogo, donde el sino-tibetano daba 97/3 contra 47/13. La razón es la aridez de la familia — con 2,75 consonantes de media hay ternarios de sobra con diez lenguas o más.

Pero el sesgo de rama es el más fuerte, y va en las dos direcciones. El bahnárico casi triplica su peso; y la rama que más códigos aporta al catálogo, la khasi-palaung, queda por debajo de la mitad del suyo. Eso no había pasado en las otras dos auditorías, donde la rama grande se inflaba y punto. La explicación está en el alcance: el khasi-palaung tiene trece lenguas con listas largas —13.469 formas, la rama de más formas de la familia— así que produce muchos códigos, pero códigos de pocas lenguas. El bahnárico tiene veintitrés lenguas y produce códigos que llegan lejos.

Y el peárico, tercera rama del catálogo con el 16,6 %, no tiene ni un código aquí.

El sesgo más grande es el que más conviene al lector, y por eso hay que decirlo primero: los códigos que cruzan cuatro ramas o más son el 10,9 % del catálogo y el 60 % de este anexo. Es la proporción más alta de la serie con diferencia. El anexo enseña lo genealógicamente interesante porque el alcance y el cruce de ramas van juntos en una familia con doce ramas repartidas — no porque se hayan elegido.

Auditoría: treinta y uno de cien dependen de la tabla

Treinta y un códigos tienen al menos una posición donde la normalización funde dos segmentos que tres lenguas o más escriben distinto. Queda entre los 9 de 91 del afroasiático y los 68 de 100 del sino-

tibetano, y casi todos son de dos clases:

1 · Lugar nasal. $t \cdot n$ | tejer es n en 20 lenguas y n en 8; $h \cdot n$ | fuego, n en 15 y n en 5; $b \cdot n$ | lleno, n en 14 y n en 4. La tabla funde n , n y n en n , y en una familia con tres nasales contrastivas eso se nota en un código de cada diez.

2 · La glotal. $n \cdot h$ | esto es $?$ en 7 y h en 5; $l \cdot h$ | hoja, igual; $t \cdot h$ | tierra, $?$ en 17 y h en 4. La tabla escribe h las dos, y en austroasiático la oclusiva glotal final es un segmento de pleno derecho.

Y hay un tercer caso, que es el interesante. $s \cdot n$ | pie reúne dz en 19 lenguas y ts en 6; $s \cdot m$ | pájaro, s en 22 y ts en 15; $c \cdot n$ | pie, c en 23 y j en 12. Son alternancias de sonoridad y de africación en la serie palatal, que es exactamente donde la fase 6 encontró su clase más fuerte — $c \leftrightarrow s$ a $16,8\times$ —. La tabla las funde *antes* de mirar y la fase 6 las redescubre *después*: el resultado coincide y el fundamento no, y esa distinción es el paper entero.

Lo que este anexo lleva impreso y el código no ve

De las 2.257 formas que citan los cien códigos, 374 (17 %) llevan al menos una marca que el esqueleto ignora, casi toda tono. Queda entre el 0 % del anexo trans-neoguineano y el 75 % del sino-tibetano, y es la cifra que hay que tener delante para saber cuánto fiarse de lo que se está viendo: en una de cada seis formas, el lector ve algo que el procedimiento no usó.

El enriquecimiento, otra vez, no ordena por antigüedad

Los diez de mayor enriquecimiento de este anexo:

Código	Concepto	Lenguas	Enriq.	Ramas
$p \cdot l \cdot m$	sanguijuela	15	$1417\times$	1
$p \cdot r \cdot t$	plátano	19	$1303\times$	1
$p \cdot l \cdot j$	fruta	17	$427\times$	1
$d \cdot m$	maduro o cocido	14	$379\times$	1
$h \cdot w$	subir	14	$301\times$	1
$m \cdot k$	tocado (prenda)	14	$277\times$	5
$r \cdot t$	comprar	12	$253\times$	1
$h \cdot r$	gallina	14	$245\times$	3
$t \cdot r \cdot n$	camino o senda	18	$243\times$	2
$s \cdot n \cdot m$	año	12	$238\times$	4

Siete de los diez están dentro de una sola rama, y es el mismo patrón que el anexo trans-neoguineano dio en estado puro: un objeto o un bicho local, con nombre largo y raro dentro del inventario de su rama, marca el máximo. El enriquecimiento mide distintividad, no profundidad.

Y en el otro extremo están los códigos que más ramas cruzan: $h \cdot n$ | yo a $18\times$ en seis ramas, $k \cdot t$ | morder a $21\times$ en seis, $s \cdot n$ | hueso a $14\times$ en cinco. Los mejores candidatos genealógicos del anexo están abajo del todo por enriquecimiento, porque están hechos de consonantes frecuentes. Las dos columnas no miden lo mismo y conviene no leerlas juntas.

Y una cosa que este anexo hace mejor que los dos anteriores

«Otros» se lleva solo 22 de los 100. En el anexo sino-tibetano eran 55 y en el trans-neoguineano 47.

No es mérito nuestro: las listas austroasiáticas de Lexibank son vocabulario básico de verdad, mientras que las sino-tibetanas traían medicina, algodón, acero y arado. Se nota en las secciones —diecisiete

códigos de partes del cuerpo, dieciséis de verbos, doce de naturaleza— y se nota en el problema del préstamo, que aquí pesa mucho menos.

Índice inverso: códigos que reaparecen en varios conceptos

Diecinueve códigos cubren cincuenta de las cien entradas, y es el índice más limpio de los tres anexos auditados: en el sino-tibetano eran veintitrés cubriendo setenta y nueve, y *s·n* se llevaba dieciséis conceptos él solo. Aquí el que más se lleva son cinco.

La tabla de abajo enseña los seis que reaparecen tres veces o más, y uno de ellos hay que leerlo despacio porque no es colisión:

s·k sale con «pelo» (40 lenguas), «pelo» otra vez (26), «pluma» (23) y «pluma o pelaje o pelo» (13). Las cuatro son el mismo campo, y dos de ellas son la misma glosa escrita con distinta caja —HAIR y hair, que el corpus trae como conceptos distintos—. No es que el esqueleto no discrimine: es que la lista de conceptos pregunta cuatro veces por lo mismo. §3 del paper cuenta cinco casos así en esta familia.

Los otros cinco sí son colisión, y de la clase manejable. *h·n* cubre yo, fuego, oler, ganso y comer con enriquecimientos de 11× a 58×; *k·t*, rana, nacer, morder y raspar. Conceptos sin relación bajo un esqueleto corto: el código agrupa por disponibilidad tanto como por historia, y el código solo no dice cuál.

Código	Conceptos con los que aparece, y su enriquecimiento
<i>h·n</i>	<i>i</i> (20, 18×) · <i>fire</i> (20, 20×) · <i>smell</i> (14, 22×) · <i>goose</i> (13, 58×) · <i>eat</i> (12, 11×)
<i>s·k</i>	HAIR (40, 82×) · <i>hair</i> (26, 75×) · <i>feather</i> (23, 51×) · <i>feather or fur or hair</i> (13, 62×)
<i>n·m</i>	<i>cry</i> (25, 89×) · <i>blood</i> (16, 30×) · <i>piss</i> (15, 66×) · <i>sweet</i> (14, 51×)
<i>k·t</i>	<i>frog</i> (20, 69×) · <i>be born</i> (20, 83×) · <i>bite</i> (14, 21×) · <i>scrape</i> (13, 92×)
<i>k·n</i>	<i>child</i> (19, 42×) · <i>boy or son</i> (17, 68×) · <i>muroid</i> (mouse or rat) (14, 27×) · <i>wide</i> (12, 30×)
<i>c·h</i>	<i>sick</i> (13, 92×) · <i>louse</i> (12, 38×) · <i>sour</i> (11, 76×)

Numerales

Cuatro códigos, y son numerales de la unidad, no compuestos de la decena como en sino-tibetano. *m·j* |uno en 31 lenguas y siete ramas, con 153× de enriquecimiento, es el mejor de la sección: es el *muay* austroasiático, y cruza más de la mitad de las ramas de la familia.

c·t |diez es el caso de la auditoría: *c* en 7 lenguas y *ɟ* en 5, una alternancia de sonoridad palatal que la tabla funde y que la fase 6 redescubre como clase.

m·j |one · binario · 31 lenguas · enriquecimiento 153× · emergente *m·1*

Alak mo:j	Brao mu:j	Chong moj.D	Cua muj
Jeh mu:j	K.Surin môj	Kasseng mu:j	Khmu mo:j
Mah Meri muj	Mlabri mɔj	Nyaheun muj	Oy muj
Rengao mɔj	Sapuan mu:j	Semaq Beri Brwe muj	Stieng muoj

b·r |two · binario · 26 lenguas · enriquecimiento 146× · emergente *2·3*

Brao ba:r	Ceq Wong ber	Chrau bar	Halang ba:r
Kasseng ba:r	Khmer Siem Reap bir	Khmu ba:r	Lao-Katu bar
Mlabri ber	MnongRlm ba:r	Nge bar	Nyakur bar.1
Pakoh bar	Santali bar	Sapuan ba:r	Stieng ba:r

p·n|four · binario · 22 lenguas · enriquecimiento 78× · emergente 2·n

Alak po:n	Brao puan	Ceq Wong pan	Chrau puon
Halang puan	Jeh puan	Kasseng puən	Mel poan
MnongRlm puon	Nyaheun puan	Palaung phon	Ramam pɔŋ
Rengao pu:n	Rumai phon	Sedang pyn	Sre puan

c·t|TEN · binario · 12 lenguas · enriquecimiento 152× · emergente 1·1

Alak c'it	BahnarPleiku ʒit	Brao cit	Cheng cit
Halang ʒə:t	JruLaven cet	Mel c.ɬ:t	Nyaheun cit
Ramam cuot	Sapuan ʒit	Sedang cat	Sre ʒət

Partes del cuerpo

Diecisiete códigos, la sección más poblada, y es donde está el techo del paper: **m·t|ojo** en 64 lenguas y once ramas, **m·h|nariz** en 58 y diez. El *mat* y el *muh*.

Merece leerse al lado el par de «pie», que §6.2 usa como su mejor candidato a isoglosa: **c·n** en 37 lenguas y ocho ramas frente a **s·n** en 37 y nueve. Mismo concepto, misma segunda consonante, y la primera reparte la familia. Con la columna emergente delante se ve además que c y s están en la misma clase, así que el par es exactamente la alternancia palatal que la fase 6 encontró por su cuenta.

m·t|eye · binario · 64 lenguas · enriquecimiento 147× · emergente m·1

Ahlao mat	Batek Deq Koh mət	Cheng mat	Cua mat
Jeh mat	Kensiw Kedah met	Ksinmul mat	Mah Meri mət
Mang Ch mat. ⁵¹	Mlabri mat	Nge mat	Oy mat
Rengao mat	Semai Kampar mat	Sre mat	Tenen Paborn mət

m·h|nose · binario · 58 lenguas · enriquecimiento 100× · emergente m·1

Alak muh	Brao muh	Chrau muh	Halang muh
Kasseng moh	Ksinmul moh	M.Bunong muh	Mel moh
Mon muh	Nyaheun muh	Parengi mu?	Prai muh
Samre muh.B	Semaq Beri Brwe muh	Sora mʔu	Taoi moh.A

s·k|HAIR · binario · 40 lenguas · enriquecimiento 82× · emergente 1·4

Alak sɔk	Bugan sak.T.55	Didra sak	Jah Hut sɔk
JruLaven sok	Kasseng sok	Kintaq sɔk	Lanoh Kertei sɔk
Mah Meri suk	MnongRlm sɔk	Nyaheun sɔk	Oy sɔk
Samre suk.A	Semai Kampar sp:k	Stieng chɔk	Taoi sɔk.A

c·n|foot · binario · 37 lenguas · enriquecimiento 61× · emergente 1·n

Ahlao ci:ŋ ¹	Batek Deq Koh can	Ceq Wong can	Chrau ʒəŋ
Jeh ʒ.ɔ:ŋ	Kensiw Kedah can	Khmer Siem Reap cə:ŋ	Maleng Bro ciŋ+
Mang Ch cuaŋ. ⁴	Mel c.ə:ŋ	Nyaheun ʒiŋ	Palaung ʒawŋ
Phong ci:ŋ. ²¹	Sapuan ʒiŋ	Sre ʒəŋ	Tampuan ciŋ

s·n|foot · binario · 37 lenguas · enriquecimiento 33× · emergente 1·n

Arem tɕiŋ	Chrau ʒəŋ	Didra ʒeŋ.L	Jah Hut ʒəŋ
K.Surin câŋ	Khmer Siem Reap ʒəŋ	M.Bunong ʒəŋ	Mlabri ʒəŋ
Mon ʒəŋ	Nyakur chuŋ.2	Pangkham tɕeŋ	Rumai ʒân.51
Sedang ceəŋ.L	Sre ʒəŋ	Tampuan cuŋ	Wa čaŋ.B

t·h|BREAST · binario · 30 lenguas · enriquecimiento 45× · emergente 1·1

Alak tɔh	BahnarPleiku tɔh	Bulang tɪh. ⁵¹	Cheng tɔh
Cua toh	Halang tuh	JruLaven toh	M.Bunong tɔh
MnongRlm tɔh	Mon tah	Nyaheun tɔh	Oy tɔh
Ramam tɔh	Sedang tuh.T	Semelai tuh	Tampuan toh

s·k|hair · binario · 26 lenguas · enriquecimiento 75× · emergente 1·4

Ahlao sɔk ¹	Alak sɔk	Ceq Wong sɔk	Cheng sɔk
Halang sɔk	Jeh suk	JruLaven sɔk	Kasseng sok
Mel sɔk	MnongRlm sɔk	Nyaheun sɔk	Phong suk
Rengao tsəŋ	Sapuan sɔk	Sedang sak	Tampuan sɔk

l·m|liver · binario · 19 lenguas · enriquecimiento 74× · emergente 3·m

Ahlao loəm.C	Arem lùəm	Chong lo:m.C	Chut loəm.2
Iduh lú:m	Kasong lɔ:m.C	Lao-Katu lɔm	Maleng Bro loəm.1
Malieng lɔm.33	Malieng lɔm.14	May lɔm ²	Nge lɔm
Pakoh lɔm	Phong lom	Samre luəm.B	Taoi lɔm.C

k·l·m|liver · ternario · 19 lenguas · enriquecimiento 136× · emergente 4·3·m

Alak klə:m	BahnarPleiku klə:m	Brao klə:m	Cheng klə:m
Chrau khləm	Halang k.l.ɛ:m	JruLaven klɔ:m	Kasseng klɔ:m
M.Bunong kləm	Mel klə:m	Nyaheun klɔ:m	Rengao k.l.ɛ:m
Sapuan klə:m	Sedang klɪam	Sre klə:m	Stieng klə:m

k·j|head · binario · 17 lenguas · enriquecimiento 50× · emergente 4·1

Batek Deq Koh kuy	Car kuj	Ceq Wong kuy	Jah Hut kɔj
Jahai Banun kuj	Kensiw Kedah kuj	Kintaq kuy	Lanoh Kertei kuy
Mah Meri koy	Menriq Lah kuy	Nancawri kúj	Semai Kampar ku:y
Semaq Beri Brwe koj	Semelai khoj	Semnam Bal ku:y	Temiar Kelantan ku:y

n·m|blood · binario · 16 lenguas · enriquecimiento 30×

Bulang nam. ⁵¹	Chrau nʰa:m	Juang ipam	Kharia epam
Kme-2 na:m. ³¹	Mel nʰa:m	Meng Dan nɪm.45	Palaung n̩a:m
Paliu nʰam. ⁵¹	Pangkham n̩a:m	Pear n̩a:m	Prai n̩iam
Rumai nham.412	Taoi n̩am.C	Wa nham.A	Xiang Zhai Tang n̩a:m

s·n|bone · binario · 16 lenguas · enriquecimiento 14× · emergente 1·n

Arem cɪeŋ	Chut sàŋ	Kharia ʒaŋ	Khasi sh'ien
Malieng saŋ.55	Malieng sapɪɪ	Mang Ch čaŋ.31	May caŋ ²
Mundari ʒaŋ	Mường [Hoa Binh] siəŋ ¹	Ngu n C	Liêm siəŋ ¹
Phong siəŋ. ²⁴	Santali ʒaŋ	Sora əʒaŋ	Paliu si. ⁵⁵ _aŋ. ⁵¹
			Vietnamese swəŋ

m·h·m|blood · ternario · 14 lenguas · enriquecimiento 220× · emergente m·1·m

Car ma-ham	Ceq Wong mhǎm	Chong mǎha:m.A	Didra mǎham
Halang mǎha:m	Kensiw Kedah mhǎm	Kintaq mhǎm	M.Bunong m'ham
Mah Meri maham	Rengao mǎha:m	Sedang mǎheam	Semaq Beri Brwe maham
Semelai maham	Tenen Paborn mahim		

k·l·k|INTESTINES · ternario · 13 lenguas · enriquecimiento 144× · emergente 4·3·4

Alak kla:k	BahnarPleiku kla:k	Brao kla:k	Cua kla:k
Halang kla:k	Jeh kla:k	JruLaven kla:k	Kasseng kla:k
Nyaheun kla:k	Ramam kla:k	Rengao kla:k	Sapuan kla:k
Tampuan kla:k			

s·m|blood · binario · 13 lenguas · enriquecimiento 28× · emergente 1·m

Chut ǎs□ m	K.Surin chiǎm	Kha Poong asam ^{33?}	Khasi snam
Khmer Siem Reap ǎham	Lyngngam snam	Maleng Bro asəm+	Malieng asam.51?
Malieng asam+1	Mon chim	Nyakur chim.1	Phong sa:m. ⁴⁴
Toum sa:m			

m·m|blood · binario · 11 lenguas · enriquecimiento 60×

Ceq Wong mhǎm	Iduh mǐm	Khmu ma:m	Ksinmul miǎm
M.Bunong mham	Mel m ^h a:m	MnongRlm mham	Parengi miam
Phong mim	Sre mham	Stieng mham	

s·h·n|bone · ternario · 11 lenguas · enriquecimiento 151× · emergente 1·1·n

Bulang sa. ³¹ _ʔaŋ. ³³	Jah Hut ʔəʔaŋ	Jahai Banun ʔəʔeŋ	K.Surin cʔʌŋ
Khasi ʃʔeŋ	Khmer Siem Reap chəʔɛŋ	Mlabri ca ʔeŋ	Phong sʔiŋ
Semaq Beri Brwe ʔaʔaŋ	Semelai ʔəʔaŋ	Wa səʔaŋ.B	

Parentesco y persona

Cuatro códigos. k·n sale con «niño» (19 lenguas) y «chico o hijo» (17) —dos glosas de un mismo campo, otra vez la lista— y también con «ratón» y «ancho», que ya no lo son.

h·n|I · binario · 20 lenguas · enriquecimiento 18× · emergente 1·n

BahnarPleiku ʔip	Bulang ʔoŋ. ³³	Ceq Wong ʔip	Chong ʔiŋ.A
K.Surin ʔap	Kasong ʔiŋ.A	Ksinmul ʔap	Menriq Lah ʔip
Phong ʔeŋ	Prai ʔəŋ	Samre ʔiŋ.A	Semai Kampar ʔen
Semelai ʔəŋ	Semnam Bal ʔi:ŋ	Sre ʔap	Tampuan ʔap

k·n|CHILD · binario · 19 lenguas · enriquecimiento 42× · emergente 4·n

Ahlao kɔ:n ¹	Arem kon	Batek Deq Koh ken	Chut kɔn
Jahai Banun ken	Khasi k ^h un	Khmu kɔ:n	Lanoh Kertei kɔn
Lyngngam k ^h on	May kɔn ¹	Menriq Lah ken	Palaung kuən
Phong kɔ:n	Rumai kɔ:n	Semnam Bal kuo:n	Toum kɔ:n

m·h|THOU · binario · 15 lenguas · enriquecimiento 27× · emergente m·1

Ahlao miʔ.B	Batek Deq Koh mɔh	Car meh	Ceq Wong mǎʔ
Jahai Banun miʔ	Kintaq muh	Kme-2 meʔ. ⁵³	Ksinmul mih
Lanoh Kertei mih	Menriq Lah mɔh	Prai mah	Semnam Bal mih
Tenen Paborn mɔh	Wa maiʔ.A	Xu meʔ.55	

n·h|THIS · binario · 12 lenguas · enriquecimiento 32× · emergente n·1

Car ɲih	Cua neʔ	JruLaven neʔ	Khmer Siem Reap nih
Khmu niʔ	Lanoh Kertei nɔɔh	Lao-Katu neʔ	Mon ɲaʔ
Mundari ni=iʔ	Pakoh nnəh.L	Semaq Beri Brwe naʔ	Semnam Bal nɔh

Naturaleza y cielo

Doce códigos. d·k|agua en 30 lenguas es el *daak* austroasiático; h·n|fuego en 20 y p·l·n|cielo en 15 completan lo esperable. Los tres dependen de la tabla en algún grado —d/ɖ en el primero, n/n en el segundo, ɲ/p en el tercero—, y los tres son códigos reales de todos modos.

d·k|water · binario · 30 lenguas · enriquecimiento 99× · emergente 1·4

Ahlao dak.C	Arem dɛək	Brao da:k	Chut da:k. ³
Halang da:k	JruLaven da:k	Kha Poong da:k. ³³	M.Bunong dak
Malieng dak.33	Malieng dakɿ1	Mel d.ɔ:k	Mường [Hoa Binh] da:k ⁷
Nyakur dak.1	Phong da:k. ²¹	Sapuan da:k	Stieng da:k

t·h|earth · binario · 21 lenguas · enriquecimiento 29× · emergente 1·1

BahnarPleiku tɛh	Batek Deq Koh teʔ	Ceq Wong tɛʔ	Cheng tɛʔ
Jah Hut teʔ	Jahai Banun teʔ	Kensiw Kedah teʔ	Lanoh Kertei tɛʔ
Mah Meri teʔ	Mel tɛh	MnongRlm tɛh	Nyakur tiʔ.1
Semai Kampar teʔ	Tampuan tɛh	Temiar Kelantan tɛʔ	Tenen Paborn tieʔ

h·n|fire · binario · 20 lenguas · enriquecimiento 20× · emergente 1·n

Alak ʔu:n	BahnarPleiku ʔun	Brao ʔun	Cheng ʔun
Halang ʔun	Jeh ʔon	JruLaven ʔun	Kasseng ʔu:n
Mel ʔon	MnongRlm ʔɿn	Nyaheun ʔun	Oy ʔun
Rengao ʔun	Sapuan ʔun	Sedang ʔon	Stieng ʔon

h·s|fire · binario · 20 lenguas · enriquecimiento 117× · emergente 1·1

Batek Deq Koh ʔɔs	Ceq Wong ʔɔs	Jah Hut ʔɔs	Jahai Banun ʔɔs
Kintaq ʔɔs	Lanoh Kertei ʔos	Lao-Katu ʔos	Mah Meri ʔus
Nge ʔujh	Pakoh ʔujh	Phong ʔos	Semai Kampar ʔo:s
Semelai ʔus	Semnam Bal ʔo:s	Sre ʔous	Temiar Kelantan ʔo:s

m·n|night · binario · 20 lenguas · enriquecimiento 33×

Alak maɲ	BahnarPleiku maɲ	Brao maɲ	Cheng maɲ
Halang maɲ	Jeh maɲ	JruLaven maɲ	Kasseng maɲ
Mel maɲ	MnongRlm maɲ	Nyaheun maɲ	Oy maɲ
Sapuan maɲ	Sedang maɲ	Sre maɲ	Stieng maɲ

t·m|stone · binario · 16 lenguas · enriquecimiento 39× · emergente 1·m

Alak tāmɔ:	BahnarPleiku tāmɔ:	Brao tāmɔ:	Chrau tāmɔ:
Halang tāmou	Jeh tāmou.T	JruLaven tāmɔ:	K.Surin thmâ
Kasseng tāmɔ	Khmer Siem Reap thmɔ:	Nge tāmɔ	Oy tāmɔ
Samre thamo.C	Sapuan tāmɔ:	Semelai tāmou	Stieng tāmou

p·l·n|sky · ternario · 15 lenguas · enriquecimiento 124× · emergente 2·3·n

BahnarPleiku plɛŋ	Chong phliŋ.A	Halang pli:ŋ	Jeh pli:ŋ
Kasong phliŋ.A	Kasseng pli:ŋ	Mang Ch pli:ŋ. ⁶	Palaung plɛŋ
Pear plɛŋ	Pusing plɛŋ	Ramam plɛŋ	Rengao plɛŋ
Rumai plân. ^{4 1 2}	Sedang plɛŋ	Tampuan plainŋ	

t·r·n|ROAD · ternario · 14 lenguas · enriquecimiento 138× · emergente 1·3·n

Chrau trɔŋ	Didra truon	Halang truon	Jeh truon.L
Kasseng truon	M.Bunong trɔn	MnongRlm trɔŋ	Nyaheun truon
Oy truon	Rengao trɔŋ.T	Sedang troan.T	Semaq Beri Brwe taruŋ
Semelai taruŋ	Stieng trɔŋ		

k·j|moon · binario · 14 lenguas · enriquecimiento 42× · emergente 4·1

BahnarPleiku kʰɛ:j	Brao kʰaj	Cheng kʰaj	Chrau khaj
Halang khaj	Iduh kâj	Jeh kʰɛ:j	JruLaven khaj
M.Bunong khaj	Mel kʰəj	Ramam kʰaj	Rengao kʰɛj
Stieng khaj	Tampuan kʰaj		

b·t·h|stone · ternario · 13 lenguas · enriquecimiento 195× · emergente 2·1·1

Batek Deq Koh batu?	Jahai Banun batu?	Kensiw Kedah batu?	Kintaq batu?
Lanoh Kertei batu?	Mah Meri batu?	Menriq Lah batu?	Semai Kampar batuh
Semaq Beri Brwe batu?	Semelai batu?	Semnam Bal batu?	Temiar Kelantan batu?
Tenen Paborn batu?			

s·m·n|star · ternario · 12 lenguas · enriquecimiento 197× · emergente 1·m·n

Bulang sa. ^{3 1} _məiŋ. ^{5 1}	Chrau səmən	Iduh samêŋ	Mel səmən
Mon saman	Palaung sameŋ	Pangkham samən	Pear s`man
Phong smeŋ	Sre səmən	Stieng səmən	Xiang Zhai Tang samən

s·n·m|year · ternario · 12 lenguas · enriquecimiento 238× · emergente 1·n·m

BahnarPleiku sənam	Cua sanim	Kha Poong sənam. ^{3 3}	Meng Dan sənam
Palaung sanəm	Pangkham sanam	Pear s`nam	Rumai sanam
Sre sənam	Stieng sənam	Tampuan sanam	Xiang Zhai Tang sanəm

Animales

Nueve códigos, y aquí están dos de los tres enriquecimientos más altos del anexo: **p·l·m|sanguijuela** a 1417× y **h·r|gallina** a 245×. El primero es de una sola rama; el segundo cruza tres.

La sanguijuela es el ejemplo perfecto de la advertencia de la cabecera: un bicho con nombre específico dentro de una rama marca el máximo de una columna que el lector interpreta como profundidad.

s·m|bird · binario · 42 lenguas · enriquecimiento 93× · emergente 1·m

Alak cem	BahnarPleiku se:m	Chrau sum	Jah Hut cem
Kasong chim.B	Khmu si:m	Lyngngam sim	Mel sim
Nge cem	Nyaheun cem	Pangkham sim	Phong som
Rengao cim.L	Sedang cem.L	Sre se:m	Vietnamese chim

c·m|bird · binario · 28 lenguas · enriquecimiento 158× · emergente 1·m

Ahlao ci:m ¹	Alak ce:m	Cheng ce:m	Chut icim.1
Iduh ciəm	Jah Hut cem	JruLaven ce:m	Kasseng ce:m
Muong [Hoa Binh] chim	Nyaheun ce:m	Ramam cem	Samre chi:m.C
Sedang cem	Semelai cim	Stieng com	Vietnamese chim

s·k|feather · binario · 23 lenguas · enriquecimiento 51× · emergente 1·4

Bulang suk. ^{3 1}	Ceq Wong sok	Chong suk.A	Halang sok
Jah Hut sok	JruLaven sok	Kasong suk.A	Kensiw Kedah sok
Kme-2 suk. ^{5 3}	Ksinmul sok	Nyaheun sok	Oy sok
Paliu suk. ^{3 3}	Phong suk. ^{4 5}	Sedang sak.L	Semai Kampar sok

k·h|fish · binario · 17 lenguas · enriquecimiento 30× · emergente 4·1

Ahlao ka?.B	Arem akeə?	Ceq Wong kiə?	Jah Hut ka?
Kha Poong aka. ^{1 3} ?	Khmu ka?	Kintaq ka?	Kme-2 a_kha?. ^{5 3}
Lanoh Kertei ka?	Mlabri ka?	Nyakur ka?.1	Paliu kha. ^{7 5 5}
Semai Kampar ka?	Semaq Beri Brwe kah	Semnam Bal ka?	Temiar Kelantan ka?

s·j|louse · binario · 15 lenguas · enriquecimiento 116× · emergente 1·1

Cua saj	JruLaven cej	K.Surin caj	Kasseng caj
Khmer Siem Reap caj	Ksinmul cej	Malieng cəj.51?	Mon caj
Nancawri séj	Nyakur caj.1	Oy caj	Sedang caj.L
Sre saj	Tampuan səj	Vietnamese tɕəj	

h·r|CHICKEN · binario · 14 lenguas · enriquecimiento 245× · emergente 1·3

Alak ʔe:r	Cheng ʔiar	Chrau ʔier	Halang ʔiər
Iduh ʔiər	JruLaven ʔiar	Kasseng ʔiər	MnongRlm ʔier
Pear ʔiar	Ramam ʔaiJ	Rengao ʔi:r	Sapuan ʔiar
Sre ʔiar	Stieng ʔier		

s·t|BEE · binario · 13 lenguas · enriquecimiento 85× · emergente 1·1

Alak su:t	BahnarPleiku su:t	Cheng sut	Chrau sɪt
Halang sut	JruLaven su:t	Kasseng su:t	Nyaheun sut
Rengao tsu:t	Sapuan sut	Sedang sot	Sre sut
Stieng sot			

h·n|GOOSE · binario · 13 lenguas · enriquecimiento 58× · emergente 1·n

Ahlao han.C	Bulang han. ^{3 3}	Chong ha:n.A	Kasong ha:n.E
Kme-2 han. ^{1 3}	Nyakur ha:ŋ	Pear han	Phong ha:n. ^{4 2}
Prai ha:n	Rumai han. ^{5 5}	Samre ha:n.A	Wa han.A
Xu han. ^{3 5}			

c·h|louse · binario · 12 lenguas · enriquecimiento 38× · emergente 1·1

Batek Deq Koh ce?	Ceq Wong ce?	Jah Hut ce?	Jahai Banun ci?
Kensiw Kedah ce?	Kintaq ce?	Lanoh Kertei ce?	Menriq Lah ce?
Semai Kampar ce:?	Semaq Beri Brwe cih	Semnam Bal ce?	Temiar Kelantan ce:?

Plantas y comida

Nueve códigos, y **p·r·t|plátano** a 1303× es el segundo mayor enriquecimiento del anexo, de una sola rama. Junto a **p·l·j|fruta** (427×) y **d·m|maduro o cocido** (379×) forman un bloque de vocabulario agrícola muy distintivo y muy local — que es lo mismo que decir que el enriquecimiento no distingue lo antiguo de lo compartido por vecindad.

r·h|root · binario · 23 lenguas · enriquecimiento 114× · emergente 3·1

Alak rɛ:h	BahnarPleiku rəh	Car rəh	Didra reh
Halang riəh	K.Surin rih	Kasong re:h.A	M.Bunong rəh
Mel əriəh	MnongRlm rəh	Nyaheun riah	Nyakur rih
Pusing riah	Rengao rih	Samre ri:h.C	Semaq Beri Brwe rəh

s·c|meat · binario · 18 lenguas · enriquecimiento 197× · emergente 1·1

Batek Deq Koh sɛc	Ceq Wong sac	Chong chuc.C	Jah Hut sɛc
Jahai Banun sɛc	Kensiw Kedah sec	Khmer Siem Reap sac	Kintaq sec
Menriq Lah sɛc	Pakoh sec.T	Prai sec	Semai Kampar sɛc
Semelai sɛc	Semnam Bal sɛc	Taoi sɛc.A	Temiar Kelantan sec

p·l·j|fruit · ternario · 17 lenguas · enriquecimiento 427× · emergente 2·3·1

Alak palaj	BahnarPleiku plɛ:j	Brao plaj	Cheng plaj
Chrau plaj	Cua plaj	Halang plaj	Jeh plaj
JruLaven plaj	Kasseng plaj	Mel pla:j	Ramam plaj
Sapuan plaj	Sedang plaj	Sre plaj	Stieng plaj

h·l|leaf · binario · 17 lenguas · enriquecimiento 72× · emergente 1·3

BahnarPleiku hla:	Chut ʔula.3	Cua hla:	Didra hla
Kasseng ʔəla:	Lao-Katu ʔala	May ʔula³	Nge hla
Nyaheun hla	Nyakur hlâ:	Oy hla	Palaung hla
Ramam hla:	Rengao hla:	Sedang hla	Tampuan hla:

b·h|salt · binario · 15 lenguas · enriquecimiento 62× · emergente 2·1

Alak bəh	BahnarPleiku bəh	Brao bəh	Cheng bəh
Cua bəh	Halang bəh	Jeh bəh	JruLaven bəh
Kasseng bəh	MnongRlm bəh	Nyaheun bəh	Rengao bəh
Sapuan bəh	Sre bəh	Stieng bəh	

r·s|root · binario · 14 lenguas · enriquecimiento 143× · emergente 3·1

Jeh riajh.T	JruLaven rias	Kasseng rias	Khmer Siem Reap riːs
Khmu riəs	Lao-Katu rias	Nge riajh	Oy riajh
Pakoh riajh.T	Phong riəs	Semaq Beri Brwe rəs	Semelai rəs
Sre rias	Taoi riajh.C		

h·n|eat · binario · 12 lenguas · enriquecimiento 11× · emergente 1·n

Ahlao ʔən	Arem ʔăn	Bulang ʔeŋ. ³³	Chut ʔan.1
Kha Poong ʔan. ³³	Maleng Bro ʔanɿ	Malieng ʔan.33	Malieng ʔăn
May ʔăn ¹	Phong ʔan	Toum ʔan	Vietnamese ăn

l·h|leaf · binario · 12 lenguas · enriquecimiento 40× · emergente 3·1

Arem uləʔ	Chong laʔ.A	Khmu ^h laʔ	Kme-2 laʔ. ⁵³
M.Bunong l'ha	Mang Ch la. ⁵¹ hɔ. ³⁵	Meng Dan ɿɛ: hi	Palaung ɿa: he:j
Parengi olaʔ	Rumai ɿa: hɔ:j	Wa lhaʔ.A	Xu laʔ.55

s·s|meat · binario · 12 lenguas · enriquecimiento 207× · emergente 1·1

Alak sac	Jah Hut sɛc	Jahai Banun sɛc	K.Surin sac
Kasseng ʒec	Kensiw Kedah sɛc	Khmer Siem Reap sac	Lao-Katu sɛc
Nge sac	Semai Kampar sɛc	Semelai sɛc	Temiar Kelantan sez

Verbos de acción y estado

Dieciséis códigos, la segunda sección más poblada, y es una diferencia grande con las dos familias anteriores: en sino-tibetano llegaba un solo verbo y en trans-neoguineano cinco. Los verbos austroasiáticos tienen consonantes suficientes para formar códigos binarios con alcance.

k·t cubre «nacer», «morder» y «raspar» con enriquecimientos de 21× a 92×, y **h·w|** subir llega a 301× en una sola rama.

k·p|bite · binario · 36 lenguas · enriquecimiento 83× · emergente 4·2

Alak kap	Batek Deq Koh kap	Car kap	Cheng kap
Didra kap	Jah Hut kap	Jeh kap	Kasseng kap
Kme-2 kap. ⁵⁵	M.Bunong kap	Menriq Lah kap	Nyaheun kap
Ramam kap	Sapuan kap	Semnam Bal kap	Stieng kap

p·r|fly_V · binario · 35 lenguas · enriquecimiento 143× · emergente 2·3

Alak par	Brao par	Cheng par	Chut parh.1
JruLaven par	Kha Poong pɿɿɿ	Maleng Bro pɿɿɿ	Mlabri par
Mundari apir	Nyakur phar.1	Pakoh par	Phong par
Pusing pɿr	Rengao par	Sre par	Tampuan par

n·m|cry · binario · 25 lenguas · enriquecimiento 89×

Alak ni:m	BahnarPleiku nəm	Cheng ni:m	Chrau ni:m
Halang niəm	Iduh níəm	JruLaven niəm	Kme-2 ná:m. ³¹
Mel nim	Mường [Hoa Binh] nhəm	Ngu n C	Liêm nam ⁴ Phong ná:m. ⁴²
Prai ném	Ramam nim	Sapuan niəm	Stieng ni:m

n·m|PISS · binario · 15 lenguas · enriquecimiento 66×

Brao no:m	Bulang nɿm. ⁵¹	Chrau no:m	Kme-2 num. ⁵⁵
Mlabri nɿm	MnongRlm no:m	Pear nɿm	Pusing nom
Rengao nɿm	Sapuan no:m	Sedang nɿm	Sre nom
Stieng no:m	Wa nɿm.B	Xu num. ⁵⁵	

k·l·h|fall · ternario · 15 lenguas · enriquecimiento 166× · emergente 4·3·1

BahnarPleiku klich	Brao kli:h	Cua kleh	Halang klich
JruLaven kleh	Kasseng kleh	Lanoh Kertei klɨ?	Nyaheun klieh
Prai khlih	Ramam klauh	Sapuan klieh	Sedang kleh
Semnam Bal klə?	Stieng kleh	Temiar Kelantan klɨ?	

k·h|vomit · binario · 15 lenguas · enriquecimiento 42× · emergente 4·1

Batek Deq Koh kə?	Ceq Wong kə?	Jahai Banun kɨ?	Kensiw Kedah kə?
Kintaq kə?	Lanoh Kertei koo?	Mah Meri ku?	Menriq Lah kə?
Paliu ku.? ⁵⁵	Semai Kampar kə:?	Semaq Beri Brwe ku?	Semelai kʰu?
Semnam Bal ko?	Temiar Kelantan ko:?	Tenen Paborn kɨ?	

k·t|bite · binario · 14 lenguas · enriquecimiento 21× · emergente 4·1

Ahlao k̚:t	Alak kiat	Bulang kiat. ³³	Cua kat
Kasong khat.A	Kha Poong ka:t. ³³	Malieng kat. ⁴³	Malieng kat+ɪ
Mon kit	Nyakur kɪt.1	Phong kat. ²¹	Samre khat.A
Toum ka:t	Wa kiat.B		

c·n|sew · binario · 14 lenguas · enriquecimiento 48× · emergente 1·n

Alak c'i:ŋ	Brao ʃiŋ	Chrau ʃiŋ	JruLaven ʃi:ŋ
Kasseng ʃi:ŋ	Mel ʃoŋ	MnongRlm ʃiŋ	Nyaheun ʃiŋ
Nyakur chí:ŋ	Palaung ʃiŋ	Pangkham cin	Sre ʃiŋ
Stieng ʃeŋ	Tampuan cɨŋ		

h·n|smell · binario · 14 lenguas · enriquecimiento 22× · emergente 1·n

Batek Deq Koh ʔɔŋ	Ceq Wong ʔɔŋ	Jah Hut ʔɔŋ	Jahai Banun ʔɔŋ
Kensiw Kedah ʔɔŋ	Kintaq ʔɔŋ	Lanoh Kertei ʔũŋ	Mah Meri hɔn
Menriq Lah ʔɔŋ	Phong hu:ń. ²⁴	Semaq Beri Brwe ʔun	Semnam Bal ʔɔ:ŋ
Temiar Kelantan ʔu:ŋ	Tenen Paborn ʔɔŋ		

h·m|give · binario · 13 lenguas · enriquecimiento 27× · emergente 1·m

Alak ʔam	BahnarPleiku ʔam	Brao ʔam	Cua ʔam
Didra ʔam	Halang ʔam	JruLaven ʔam	Nyaheun ʔəm
Oy ʔam	Ramam ʔam	Rengao ʔam	Sapuan ʔam
Sedang ʔəm			

h·k|give · binario · 13 lenguas · enriquecimiento 30× · emergente 1·4

Batek Deq Koh ʔɛk	Ceq Wong ʔak	Jah Hut ʔɔk	Jahai Banun ʔɛk
Kensiw Kedah ʔɛk	Kintaq ʔɛk	Lanoh Kertei ʔɔok	Menriq Lah ʔɛk
Samre ʔuək.C	Semai Kampar ʔɔ:k	Semnam Bal ʔo:k	Temiar Kelantan ʔo:k
Tenen Paborn ʔɛk			

p·l|fly_v · binario · 12 lenguas · enriquecimiento 45× · emergente 2·3

Arem pāl	Chut pəl	Cua pəl	Kha Poong pǎl ²³
Ksinmul pal	Malieng pǎl	May pǎl ¹	Meng Dan pil.55
Mường [Hoa Binh] pal ¹	Phong pil	Sapuan pal	Toum pəl

c·p|WALK · binario · 11 lenguas · enriquecimiento 70× · emergente 1·2

Batek Deq Koh cip	Ceq Wong cip	Jah Hut cip	Jahai Banun cip
Kensiw Kedah cip	Kintaq cip	Lanoh Kertei cip	Menriq Lah cip
Semai Kampar ci:p	Semnam Bal ci:p	Temiar Kelantan ci:p	

n·p|count · binario · 11 lenguas · enriquecimiento 231× · emergente n·2

Ahlao năp	Chong nap.C	Iduh na ² ɔp	Kasong nap.E
Khmu nap	Mlabri nap	Nyakur nap	Phong nap. ⁴⁴
Prai nap	Pusing nap	Samre nap.B	

j·m|cry · binario · 11 lenguas · enriquecimiento 149× · emergente 1·m

Bulang jam. ⁵¹	Chong ja:m.C	Chut jam	Kasong ja:m.C
Khmu ja:m	Nyakur jâ:m	Rumai jam. ⁴¹²	Samre ja:m.B
Toum ja:m	Wa jiam.B	Xu jam. ³¹	

s·t|die · binario · 11 lenguas · enriquecimiento 32× · emergente 1·1

Alak si:t	Arem tci:t	Chrau cit	Cua se:t
Kasseng cit	Lao-Katu cet	Mang Ch sit. ⁵⁵	Pakoh cet.T
Sre chət	Stieng chet	Vietnamese ch□ t	

Objetos y cultura

Un solo código con esta etiqueta, **t·r·n|camino** o **senda**, en 18 lenguas y dos ramas a 243×. La cultura material pesa poco en este anexo, y eso es una propiedad de las listas de origen: son vocabulario básico.

t·r·n|PATH OR ROAD · ternario · 18 lenguas · enriquecimiento 243× · emergente 1·3·n

BahnarPleiku trɔ:ŋ	Brao tru:ŋ	Cheng tru:ŋ	Chrau trɔ:ŋ
Halang trɔaŋ	Jeh truəŋ	JruLaven truon	Kasseng truəŋ
MnongRlm trɔ:ŋ	Nyaheun truon	Rengao trɔ:ŋ	Sapuan tru:ŋ
Sedang trɔəŋ	Semaq Beri Brwe tɔon	Semelai tron	Stieng trɔ:ŋ

Cualidades

Seis códigos, con enriquecimientos medios y alcance corto. Es la sección donde el instrumento tiene menos que decir en todas las familias de la serie, y aquí también.

b·n|full · binario · 19 lenguas · enriquecimiento 62× · emergente 2·n

Ahlao bɔ̃n	Brao biŋ	Chrau beŋ	Halang bi:ŋ
Jeh biŋ	JruLaven bi:ŋ	Khmer Siem Reap bəŋ	Lao-Katu biŋ
M.Bunong beŋ	Mel b.ɛ:.ŋ	MnongRlm bi:ŋ	Oy biŋ
Pakoh biŋ	Rengao biŋ	Sapuan biŋ	Sre be:ŋ

t·n|bitter · binario · 16 lenguas · enriquecimiento 31× · emergente 1·n

Ahlao taŋ ⁴	BahnarPleiku taŋ	Chrau taŋ	Chut ət□ □ ŋ
Kha Poong tăŋ ^{3 3 7}	Maleng Bro táŋ†	Malieng taŋ ³	May tăŋ ³
Mel taŋ	MnongRlm taŋ	Muong [Hoa Binh] taŋ ³	Ngu□ n C□ Liêm tăŋ ³
Phong taŋ	Stieng taŋ	Tampuan taŋ	Toum taŋ

n·m|sweet · binario · 14 lenguas · enriquecimiento 51×

Alak ɲa:m	BahnarPleiku ɲa:m	Brao ɲa:m	Chong ɲa:m.D
Cua ɲi:m	Kasong ɲa:m.D	Kasseng ɲa:m	Pear ɲam
Ramam ɲa:m	Rengao ɲa:m	Rumai ɲam. ^{4 1 2}	Samre ɲa:m.C
Sapuan ɲa:m	Sedang ɲɛəm		

b·k|white · binario · 13 lenguas · enriquecimiento 93× · emergente 2·4

Brao bə:k	Cheng bə:k	Kasseng bək	Lao–Katu bək
Meng Dan buk. ³⁵	MnongRlm buok	Nge bək	Nyaheun bə:k
Oy bək	Pusing buk	Sapuan bə:k	Stieng bə:k
Tampuan ba:k			

k·r·h|old · ternario · 12 lenguas · enriquecimiento 190× · emergente 4·3·1

Alak kra?	BahnarPleiku kra?	Brao kra?	Cheng kra?
Cua kara?	Halang kra?	JruLaven kra?	Kasseng kra?
Nyaheun kra?	Rengao kra?	Sapuan kra?	Tampuan kra?

k·n|wide · binario · 12 lenguas · enriquecimiento 30× · emergente 4·n

Ahlao kŋaŋ	Chong kwa:ŋ.A	Iduh kwâ:ŋ	Kasong kwa:ŋ.E
Kha Poong kwan ^{3 3 7}	Maleng Bro kwâ:ŋ†	Mlabri kwan	Nyakur kwâ:ŋ
Phong kwa:ŋ. ^{4 4}	Prai khwa:ŋ	Pusing kuan	Samre kwa:ŋ.B

Otros

Veintidós códigos, menos de la cuarta parte del anexo. En el sino-tibetano eran cincuenta y cinco y en el trans-neoguineano cuarenta y siete.

No es una categoría residual pero tampoco es el vertedero que fue en las otras dos: son conceptos que la lista temática —escrita para vocabulario básico tipo Swadesh— no contempla por su etiqueta, no por su naturaleza. **m·k|tocado** (prenda) a 277× en cinco ramas y **s·n·m|año** a 238× en cuatro son los dos mejores, y los dos son códigos de alcance amplio que cruzan ramas: exactamente lo contrario del «Otros» de las dos familias anteriores, donde el techo era un préstamo del siglo XVII y una lista de objetos de una sola rama.

t·n|WEAVE · binario · 32 lenguas · enriquecimiento 78× · emergente 1·n

Alak ta:ŋ	Brao ta:ŋ	Ceq Wong teŋ	Chong tha:ń.A
Halang ta:n	JruLaven ta:ŋ	Kasseng ta:ŋ	Mang Ch ta:n. ⁶
MnongRlm ta:ŋ	Nyakur tâ:ñ	Phong ta:ń. ^{2 4}	Ramam taŋ
Samre tha:ń.A	Sedang teŋ	Stieng ta:ŋ	Wa taiŋ.A

p·n|SHOOT · binario · 29 lenguas · enriquecimiento 58× · emergente 2·n

Alak paŋ	Brao paŋ	Cheng paŋ	Halang pəŋ
Jeh pəŋ	Kasseng paŋ	Khmu pín	Mel paŋ
Nyaheun paŋ	Pear pɔn	Prai pheŋ	Pusing pəŋ
Rengao pəŋ	Sedang pəŋ	Stieng paŋ	Wa puŋ.A

k·t|FROG · binario · 20 lenguas · enriquecimiento 69× · emergente 4·1

Alak ki:t	BahnarPleiku kit	Brao ki:t	Cheng ki:t
Halang ki:t	Jeh kɪat	Kasseng ki:t	Kha Poong kɔ:t. ³³
MnongRlm kɪt	Nyaheun kiet	Paliu a. ³³ _khet. ³³	Ramam kɔt
Sedang kɛt	Sre kit	Stieng kɪt	Tampuan kɛt

k·t|BE BORN · binario · 20 lenguas · enriquecimiento 83× · emergente 4·1

Alak kə:t	Brao kə:t	Bulang kət. ³³	Chong kə:t.A
Iduh kə:t	JruLaven kə:t	Kasong kə:t.E	Kme-2 kət. ¹³
Nyaheun kə:t	Nyakur kə:t	Paliu kət. ⁵⁵	Pear kət
Pusing kɪt	Rumai kɪt. ⁵⁵	Samre kə:t.A	Stieng kə:t

p·r·t|BANANA · ternario · 19 lenguas · enriquecimiento 1303× · emergente 2·3·1

Alak pri:t	BahnarPleiku pri:t	Brao pri:t	Cheng pri:t
Chrau pri:t	Cua pare:t	Jeh pri:t	JruLaven pri:t
Kasseng pri:t	Mel pri:t	MnongRlm pri:t	Rengao pret
Sapuan pri:t	Sedang pri:t	Sre pri:t	Stieng pri:t

k·n|BOY OR SON · binario · 17 lenguas · enriquecimiento 68× · emergente 4·n

Alak kɔ:n	BahnarPleiku kɔ:n	Brao kuan	Chrau kɔ:n
Halang koan	Jeh kɔ:n	JruLaven kuan	Mel kɔ:n
MnongRlm kuon	Nyaheun kuan	Ramam kɔ:n	Rengao kɔ:n
Sapuan kuan	Sedang kɔn	Sre kon	Stieng kɔ:n

p·l·m|LEECH · ternario · 15 lenguas · enriquecimiento 1417× · emergente 2·3·m

Alak plə:m	BahnarPleiku plə:m	Brao plə:m	Chrau plə:m
Halang p.l.ɛ:m	Jeh p.l.ɛ:m	JruLaven plɔ:m	Kasseng plɔ:m
MnongRlm plə:m	Nyaheun plə:m	Rengao p.l.ɛ:m	Sedang pliəm
Sre plɔm	Stieng plə:m	Tampuan pləəm	

m·k|HEADGEAR · binario · 14 lenguas · enriquecimiento 277× · emergente m·4

Ahlao muok	Bulang mhɔk. ⁵¹	Chong muak.A	Kasong muək.D
Kha Poong muok. ³³	Kme-2 mok. ¹³	Mang Ch mɔk. ⁵⁵	Mlabri muak
Paliu m ¹ uək. ⁵¹	Phong muak. ⁴²	Prai muək	Rumai mhok. ⁵⁵
Samre muək.A	Wa mhok.A		

k·n|MURDID (MOUSE OR RAT) · binario · 14 lenguas · enriquecimiento 27× · emergente 4·n

BahnarPleiku kəɛ:	Brao kəɛ:	Cheng kne:	Chong khɔ:n.B
Chrau kəɛ:	Cua kane:	Halang kəɛ:	JruLaven kne:
Kasong khɔ:n.B	Mel kəɛ:	Phong kəɛ. ⁴⁴	Rengao kəni:
Samre khuən.C	Wa kiaŋ.B		

b·k|tie · binario · 14 lenguas · enriquecimiento 141× · emergente 2·4

Brao bɨk	Cheng bɨk	JruLaven bɨk	Kasseng bɨk
Lanoh Kertei bɨk	Mah Meri bək	Nyaheun bɨk	Rumai bak. ^{5 5}
Sapuan bɨk	Semai Kampar bək	Semaq Beri Brwe bək	Semelai bək
Semnam Bal bə:k	Temiar Kelantan bək		

h·w|GO UP (ASCEND) · binario · 14 lenguas · enriquecimiento 301× · emergente 1·w

BahnarPleiku ha:w	Brao ha:w	Chrau ha:w	Halang ha:w
Jeh ha:w	JruLaven haw	Mel ha:w	MnongRlm ha:w
Nyaheun ha:w	Ramam ha:w	Rengao ha:w	Sre ha:w
Stieng ha:w	Tampuan ha:w		

d·m|RIPE OR COOKED · binario · 14 lenguas · enriquecimiento 379× · emergente 1·m

BahnarPleiku du:m	Brao dum	Cheng dum	Chrau dum
Halang du:m	Jeh dɯm	JruLaven du:m	Kasseng do:m
Mel dɔm	MnongRlm dɨ:m	Nyaheun dum	Sapuan dɨm
Sre dum	Stieng dum		

h·h|WE (INCLUSIVE) · binario · 14 lenguas · enriquecimiento 93× · emergente 1·1

Batek Deq Koh he?	Ceq Wong hɛ?	Jahai Banun he?	Kensiw Kedah he?
Kintaq he?	Kme-2 ʔɛʔ. ^{5 3}	Lanoh Kertei ʔɛ?	Mang Ch ʔa. ^{3 1} _hi. ^{3 1}
Menriq Lah hi?	Semai Kampar he:ʔ	Semaq Beri Brwe heh	Semnam Bal ʔɛ:ʔ
Temiar Kelantan ʔɛ:ʔ	Wa ʔeʔ.A		

s·k|FEATHER OR FUR OR HAIR · binario · 13 lenguas · enriquecimiento 62× · emergente 1·4

BahnarPleiku sɔk	Cheng sɔk	Chut ʊsúk	Halang sɔk
Jahai Banun sɔk	Kasseng sok	Kha Poong usukɿɿ	Maleng Bro usɔk
Malieng uʂɔkɿɿ	Mel sɔk	Ramam tsak	Rengao tsɔk
Sedang sak			

c·h|SICK · binario · 13 lenguas · enriquecimiento 92× · emergente 1·1

Alak c'ɨ?	BahnarPleiku ʒɨ?	Brao ʒɨ?	Cheng ʒɨ?
Cua ʒɨ?	Halang ʒɨ?	Jeh ʒɨ?	JruLaven ʒɨ?
Nyaheun ʒɨ?	Pusing cō?	Ramam cu?	Rengao ʒɨ?
Sapuan ʒɨ?			

k·t|SCRAPE · binario · 13 lenguas · enriquecimiento 92× · emergente 4·1

Bulang khut. ^{5 1}	Chong khu:t.A	Kasong khu:t.B	Kme-2 khut. ^{1 3}
Nyakur khu:t	Paliu khɨt. ^{5 1}	Pear kat	Phong khu:t. ^{2 1}
Prai khu:t	Pusing kɔt	Rumai khiet. ^{5 5}	Wa kɔt.A
Xu kuat. ^{3 3}			

r·t|BUY · binario · 12 lenguas · enriquecimiento 253× · emergente 3·1

Alak ro:t	BahnarPleiku rɔt	Brao ruat	Cua rɔ:t
Halang ruat	Jeh ruat	JruLaven ruat	Mel ruət
MnongRlm ruot	Nyaheun ruat	Rengao r.ɔ:t	Sre roat

l·k|CHOOSE · binario · 12 lenguas · enriquecimiento 180× · emergente 3·4

Bulang lhək. ³³	Chong liək.C	Kasong liək.D	Kha Poong ali:k. ³³
Kme-2 lək. ¹³	Nyakur lə:k	Phong liək. ²¹	Prai liək
Pusing liək	Samre liək.C	Wa laik.B	Xu lək. ³³

m·m|IRON · binario · 12 lenguas · enriquecimiento 165×

Alak ma:m	BahnarPleiku ma:m	Brao ma:m	Cheng ma:m
Halang ma:m	Jeh ma:m	JruLaven mə:m	Kasseng ma:m
Nyaheun ma:m	Rengao ma:m	Sapuan ma:m	Sedang mɛəm

t·c|SELL · binario · 12 lenguas · enriquecimiento 194× · emergente 1·1

BahnarPleiku tɛc	Chrau tac	Cua tɛc	JruLaven tec
Kasseng tac	Mel tac	MnongRlm tac	Nyaheun tec
Prai thec	Sapuan tɛc	Sre tac	Stieng tac

k·c·t|kill · ternario · 12 lenguas · enriquecimiento 169× · emergente 4·1·1

Ahlao kacâ:t	Brao kəce:t	Cheng kce:t	Chut kacit
JruLaven kce:t	Lao-Katu kəcet	Malieng kacet. ²⁴	Malieng kacét
Nyakur kacét	Oy kəcet	Pakoh kəcet.T	Sapuan kacət

c·h|SOUR · binario · 11 lenguas · enriquecimiento 76× · emergente 1·1

Ahlao cu?.B	Cheng ʒu?	Chong cho?.C	Cua ʒua?
Halang ʒuə?	Iduh cɔʔɛ?	Jeh ʒua?	Kasseng ʒɔ?
Kha Poong cɔ. ³³ ?	Rengao ʒɔ?	Sapuan ʒuo?	